COMPUTER METHOD AND APPARATUS FOR DETERMINING CONTENT TYPES OF WEB PAGES

ABSTRACT OF THE DISCLOSURE

Computer method and apparatus determines content type of contents of a subject Web page. A predefined set of potential content types is first provided. For each potential content type, there are one or more tests having test results that enable quantitative evaluation of the contents of the subject Web page. A respective probability of each potential content type being detected in some contents of the subject Web page is determined. A Bayesian network combines the test results to provide indications of the types of contents detected on the subject Web page. A confidence level per detected content type is also provided. A database stores the determined probabilities and confidence levels, and thus provides a cross reference between Web pages and respective content types of contents found on the Web pages.